SID 18

```
RESULT    2
BOVANGA
LOCUS       BOVANGA       924 bp    DNA             MAM       25-NOV-1994
DEFINITION  Bos taurus (Aberdeen Angus strain) DNA fragment.
ACCESSION   L27712
VERSION     L27712.1  GI:443737
KEYWORDS    .
SOURCE      Mitochondrion Bos taurus (strain Aberdeen Angus, organelle
            Mitochondrion Bos taurus) Male DNA.
  ORGANISM  Mitochondrion Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   Loftus,R.T., MacHugh,D.E., Bradley,D.G., Sharp,P.M. and
            Cunningham,P.
  TITLE     Evidence for two independent domestications of cattle
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (7), 2757-2761 (1994)
  MEDLINE   94195822
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Bos taurus"
                     /organelle="mitochondrion"
                     /strain="Aberdeen Angus"
                     /db_xref="taxon:9913"
                     /sex="Male"
BASE COUNT      303 a    224 c    127 g    270 t
ORIGIN


  Query Match             100.0%;  Score 158;  DB 3;  Length 924;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-24;
  Matches  158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gtcaatggtcacaggacataaattatattatatatccccccttcataaaaattcccccct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745 GTCAATGGTCACAGGACATAAATTATATTATATATCCCCCCTTCATAAAAATTCCCCCCT 804

Qy     61 taaatatctaccaccacttttaacagacttttccctagatacttatttaaatttttcacg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805 TAAATATCTACCACCACTTTTAACAGACTTTTCCCTAGATACTTATTTAAATTTTTCACG 864

Qy    121 ctttcaatactcaatttagcactccaaacaaagtcaat 158
          ||||||||||||||||||||||||||||||||||||||
Db    865 CTTTCAATACTCAATTTAGCACTCCAAACAAAGTCAAT 902
```

addition, determination of the sizes and patterns of induction of the RNA transcripts was necessary to enable estimation of the redundancy of the clones. Therefore, Northern blot analysis was performed with RNA isolated from human IL-2R-positive T-cell blasts stimulated with either CHX or IL-2 alone, or with a combination of the two agents.

Hybridization of the RNA with probes generated from the inserts of each of the 18 putative clones resulted in the identification of 4 clones that were solely CHX-induced. For the remaining 14 clones, the induction by the combination of IL-2 and CHX could not be accounted for by the effects of CHX alone. Based upon the patterns of induction and approximate sizes of the RNA transcripts, 8 readily distinguishable and apparently unique IL-2-induced genes were discerned, as partial sequences, among these 14. These are described in Table X.

TABLE 2

Clone	Nucleotide Sequence	Insert (kb) Size of Partial Sequence	RNA (bases) Size of Partial Sequence	IL-2 Induction
CR1 (1A8)	nucleotide 857 to 2406 of SEQ ID NO: 1	1.6	2406	24
CR2 (1F5)	nucleotides 1 to 163 and nucleotides 1093 to 1283 of SEQ ID NO: 3	1.1	1283	7
CR3 (10A8)	nucleotides 718 to 901 and nucleotides 2265 to 2450 of SEQ ID NO: 5	2.0	2450	22
CR4 (10D6)	nucleotides 2101 to 2291 and nucleotides 2679 to 2928 of SEQ ID NO: 7	1.0	2946	6
CR5 (10F9)	nucleotides 763 to 902 and nucleotides 1641 to 2020 of SEQ ID NO: 9	1.4	2020	>50
CR6 (11B2)	nucleotides 310 to 513 and nucleotides 687 to 1066 of SEQ ID NO: 11	1.0	1066	5
CR7 (11E6)	corresponds to nucleotides of pim-1 sequence in Selten et al.	0.7	2400	17
CR8 (13E2)	nucleotides 1721 to 1915 of SEQ ID NO: 13	1.5	2980	7

The original designations of the CR clones are included in parentheses in the left-hand column of Table II. The original designations are used herein to refer to the partial sequences shown in the column second from the left in Table II. As shown in Table II and in FIGS. 8A–8H, three of the genes, CR1, CR3, and CR5, were induced by IL-2 alone, while five of the genes, CR2, CR4, CR6, CR7, and CR8, were induced by both CHX and IL-2. In several instances, the combination of IL-2 and CHX resulted in a marked synergistic induction.

Example III

Kinetic Analysis of IL-2-Induced Gene Expression

The temporal expression of the novel, IL-2-induced genes was determined by Northern blot analysis, using RNA isolated from human IL-2R-positive T-cell blasts after IL-2 stimulation in the presence or absence of CHX. Northern blots were prepared with 15 mg total RNA isolated from G_0/G_1-synchronized human T-cells stimulated for 0, 0.5, 1, 2, 4, or 8 hours with 1 nM IL-2 or IL-2+10 mg/ml CHX. Filters were probed with the cDNA inserts of the IL-2-induced clones.

As shown in FIGS. 9A–9H, two of the genes, 1A8 (FIG. 9A) and 10D6 (FIG. 9B), exhibited rapid induction, reaching peak levels within 1–4 hr of IL-2 stimulation and returning to basal levels after 8 hr, while the other six clones (FIGS. 9C–9H) remained at elevated levels for at least 8 hr after IL-2 treatment. The magnitude of IL-2 induction of steady state RNA levels of the clones ranged from an approximately 5-fold elevation of clone 11B2 (FIG. 9F) to a greater than 50-fold stimulation of clone 10F9 (FIG. 9E) during the interval examined. These results are also summarized in Table II. Several of the clones were superinduced by CHX, with an increase observed in both the magnitude and duration of the IL-2 response.

The kinetics of induction of previously characterized IL-2-responsive genes have been found to range from those such as c-fos, which are rapidly and transiently induced within minutes of IL-2 stimulation (Dautry et al. (1988) *J. Biol. Chem.* 263:17615–17620), to those which remain at elevated levels through G_1 to S phase entry (Sabath et al. (1990) *J. Biol. Chem.* 265:12671–12678).

Example IV

Sequence Analysis of Clones Containing Ligand-Induced Genes

To verify the redundancy of the clones as estimated from Northern analysis, as well as to determine the identities of the genes, the cDNA clones were subjected to sequence analysis.

Plasmids were isolated from the clones of interest essentially as described by Kraft et al. ((1988) *Biotechniques* 6:544–547), and vector primers were used to sequence the termini of the cDNA inserts, employing the Sequenase 2.0 dideoxy sequencing kit (United States Biochemical, Cleveland, Ohio). Approximately 200 bases of sequence were attained from each end of the inserts. These partial sequences are described in Table II. Searches of the GenBank and EMBL data bases were performed with the FASTA program as described by Pearson et al. ((1988) *Proc. Natl. Acad. Sci.* (USA) 85:2444–2448).

The combination of sequence and Northern analyses revealed that the 14 putative IL-2-induced clones consisted of 8 unique genes, three of which, 1A8, 11B2, and 13E2, were isolated three times each. Searches of the GenBank and EMBL data bases with the partial sequences enabled the identification of one clone, 11E6, as pim-1, a previously characterized IL-2-induced gene (Dautry et al. (1988) *J. Biol. Chem.* 263:17615–17620; and Kakut-Houri et al. (1987) *Gene* 54:105–111) which encodes a 33 kD cytoplasmic kinase (Telerman et al. (1988) *Mol. Cell. Biol.* 8:1498–1503).

Thus, by utilizing the method of the invention seven unique IL-2 induced genes were cloned, representing novel human genes. These clones were identified after screening only approximately 800 library colonies, and thus, it is estimated that as many as 80 additional novel IL-2-induced genes remain to be detected in the 10,000-clone library.

To determine the complete sequences of these clones described in Table II the original partial cDNAs were used as probes to screen a second cDNA library. It is standard procedure to use partial cDNA inserts identified by an initial screen of a cDNA library to make radiolabeled cDNA probes to screen a second library to obtain clones with the portions missing in the initial cDNA clones. This was done, briefly, as follows: a second cDNA library was prepared from mRNA obtained from human T cells stimulated for two hours with interleukin-2 in the presence of cycloheximide by cloning into the λgt-10 phage vector using standard meth-

SID 17

```
RESULT    4
BOVHERB
LOCUS       BOVHERB       893 bp    DNA            MAM       25-NOV-1994
DEFINITION  Bos taurus (Hereford strain) DNA fragment.
ACCESSION   L27725
VERSION     L27725.1  GI:443748
KEYWORDS    .
SOURCE      Mitochondrion Bos taurus (strain Hereford, organelle Mitochondrion
            Bos taurus) male DNA.
  ORGANISM  Mitochondrion Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   Loftus,R.T., MacHugh,D.E., Bradley,D.G., Sharp,P.M. and
            Cunningham,P.
  TITLE     Evidence for two independent domestications of cattle
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (7), 2757-2761 (1994)
  MEDLINE   94195822
FEATURES             Location/Qualifiers
     source          1..893
                     /organism="Bos taurus"
                     /organelle="mitochondrion"
                     /strain="Hereford"
                     /db_xref="taxon:9913"
                     /sex="male"
BASE COUNT      292 a    216 c    125 g    260 t
ORIGIN


  Query Match             100.0%;  Score 265;  DB 3;  Length 893;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-71;
  Matches  265;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 atccaatgaattttaccaggcatctggttctttcttcagggccatctcatctaaaacggt 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      441 ATCCAATGAATTTTACCAGGCATCTGGTTCTTTCTTCAGGGCCATCTCATCTAAAACGGT 500

Qy       61 ccattctttcctcttaaataagacatctcgatggactaatggctaatcagcccatgctca 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      501 CCATTCTTTCCTCTTAAATAAGACATCTCGATGGACTAATGGCTAATCAGCCCATGCTCA 560

Qy      121 cacataactgtgctgtcatacatttggtatttttttattttggggggatgcttggactcag 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      561 CACATAACTGTGCTGTCATACATTTGGTATTTTTTTATTTTGGGGGATGCTTGGACTCAG 620

Qy      181 ctatggccgtcaaaggccctgacccggagcatctattgtagctggacttaactgcatctt 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      621 CTATGGCCGTCAAAGGCCCTGACCCGGAGCATCTATTGTAGCTGGACTTAACTGCATCTT 680

Qy      241 gagcaccagcataatgataagcatg 265
            |||||||||||||||||||||||||
Db      681 GAGCACCAGCATAATGATAAGCATG 705
```

ods. (Sambrook, J. et al. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, New York, 1989) pp. 2.82–2.122).

This second cDNA library was then screened using as probes each of the cDNA fragments obtained from the first, thiol-selected cDNA library. Candidate clones that corresponded to the correct size according to the mRNA were then subcloned and sequenced. The complete cDNA sequences (and the predicted amino acid sequences) of seven out of eight of these clones are set forth in SEQ ID NOs:1–14 and FIGS. 1–7. The complete cDNA sequence (and the predicted amino acid sequence) of the eighth clone was determined to be identical to that of the IL-2 induced gene pim-1. The nucleotide sequence as well as the predicted amino acid sequence of pin-1 are set forth at page 605 in Selten, G. et al. (1986) *Cell* 46:603-611.

Example V

Determination of Sensitivity of IL-2-Induced Gene Expression

As a further means of characterizing the regulation of expression of these genes, the sensitivity of induction to the known IL-2 functional antagonist was investigated. Human IL-2R-positive T-cell blasts were stimulated with IL-2 in the absence or presence of 0.5 mM dibutyryl-cAMP, a concentration of the membrane-permeant cAMP analog sufficient to inhibit IL-2-mediated GI progression without adversely affecting cellular viability. The effect of an equivalent molar amount of sodium butyrate, which does not inhibit the IL-2 response, was also tested to control for the actions of free butyric acid.

Northern blots were prepared as follows: Human IL-2R-positive T cells were treated with 1 nM IL-2 alone or in combination with 0.5 mM dibutyryl cAMP or sodium butyrate (NaBt) for 1, 2, or 4 hours. Filters were prepared with 15 mg total RNA and hybridized with cDNA inserts or the IL-2 induced clones.

These analyses demonstrate that the IL-2 induction of one gene, 1A8 (FIG. **10A**) is markedly inhibited when the intracellular level of cAMP is raised by the addition of dibutyryl cAMP, whereas the expression of two others, 10D6 (FIG. **10B**) and 13E2 (FIG. **10C**), is augmented approximately 3-fold. By comparison, the expression of five of the genes was not affected by elevated cAMP (FIGS. **10D–10H**). Thus, the sequences in clone 1A8 may be involved in T-cell proliferation. The fact that not all genes were sensitive to cAMP indicated that the observed results were not due to nonspecific effects, and furthermore that the previously documented down-regulation of IL-2R binding capacity by cAMP (Johnson et al. (1990) *J. Immunol.* 145:1144–1151) could not account for the inhibition of gene expression.

Example VI

Determination of Role of T-cell Receptor Activation in the Stimulation of Expression of IL-2-Induced Genes

In order to determine if activation of the T-cell receptor mediates the stimulation of expression of cytokine IL-2-induced genes, the following study was performed. Northern blots were prepared from 20 mg total cellular RNA isolated from human peripheral blood mononuclear cells (PBMCs) stimulated with a monoclonal antibody (OKT3) specific to the CD3 component of the T-cell antigen receptor complex. Blots were probed with cDNA inserts of the IL-2-induced clones. Data was determined as the mean±SEM (n=6).

By isolation of RNA at early time intervals, it was possible to identify those genes which were induced by T-cell receptor triggering in the absence of IL-2 effects. As shown in FIGS. **11A–11H**, only one of the genes, 10D6 exhibited heightened levels of expression after 2 hr of T-cell receptor activation, while the seven others were apparently insensitive to this stimulus. Two of the clones, 1F5 and 11 B2, were undetectable, even after seven days of autoradiographic exposure of the Northern blots. Two other genes, 11E6 and 13E2, were expressed at relatively high levels regardless of the stimulus; activation with anti-CD3 did not induce RNA expression beyond the level observed by culture in medium alone. Identical results were obtained after 1 and 4 hr of stimulation.

To determine whether the cells were actually activated via CD3, aliquots of the cells were left in culture for 52 hr in the presence of 10 mg/ml CHX, alone, OKT3 alone, or OKT3+CHX, after which cell cycle progression was monitored by [^{3}H]-thymidine incorporation into RNA.

As shown in FIG. **12**, the cells were sufficiently stimulated by anti-CD3. Thus, the T-cell receptor-induced expression of only one of the genes was comparable to that seen with IL-2 stimulation, while the expression of the seven others was unique to the IL-2 signaling pathway. Thus, the methods described herein to identify IL-2-induced gene successfully selected and enriched for these genes that are highly specific for cytokine (IL-2) activation.

Of the 8 IL-2 induced G_1 progression genes reported here, only one appears to also be induced during the T cell receptor-mediated competence phase of the cell cycle. Thus, while several genes such as c-fos, c-myc and c-raf-1 are known to be induced during both the initial G_0-G_1 and subsequent G_1-S phase transitions, the expression of a number of IL-2-stimulated genes is unique to the latter event. In addition, the immediate-early genes reported here appear to define a class distinct from the IL-2-induced genes isolated by Sabath et al. ((1990) *J. Biol. Chem.* 265:12671–12678). These investigators utilized a differential screening procedure to isolate genes expressed at the G_1/S phase boundary in a murine T helper clone which was stimulated with IL-2 for 20 hr in the absence of protein synthesis inhibitors. In this case, the expression of only 3 of the 21 clones isolated was inhibited by CHX, while the remainder were insensitive to this agent. This pattern of regulation markedly contrasts with the CHX superinduction observed with the immediate-early IL-2-induced genes described here. Moreover, these observations indicate that IL-2 stimulates a complex program of gene expression, ranging from those genes induced very early in G_1 through those subsequently expressed at the G_1/S phase transition.

Example VII

Cloning and Analysis of CR8

As described above, the CR8 gene encodes a novel basic helix-loop-helix (bHLH) protein. While the CR8 transcript is ubiquitously expressed in many tissues, it is induced by IL-2 as well as by IL-3 in cytokine-dependent lymphoid cell lines. In an IL-2-dependent human T cell line Kit 225, the CR8 transcript is induced not only by IL-2, but also by interferon b and forskolin, which elevates intracellular cAMP. The bHLH domain of CR8 shows the highest structural homology to a Drosophila transcriptional repressor hairy. The recombinant CR8 protein binds preferentially to

Seq ID 8

```
RESULT     8
BOVHERB
LOCUS          BOVHERB         893 bp    DNA              MAM        25-NOV-1994
DEFINITION     Bos taurus (Hereford strain) DNA fragment.
ACCESSION      L27725
VERSION        L27725.1  GI:443748
KEYWORDS       .
SOURCE         Mitochondrion Bos taurus (strain Hereford, organelle Mitochondrion
               Bos taurus) male DNA.
  ORGANISM     Mitochondrion Bos taurus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
               Bovidae; Bovinae; Bos.
REFERENCE      1  (bases 1 to 893)
  AUTHORS      Loftus,R.T., MacHugh,D.E., Bradley,D.G., Sharp,P.M. and
               Cunningham,P.
  TITLE        Evidence for two independent domestications of cattle
  JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 91 (7), 2757-2761 (1994)
  MEDLINE      94195.2
FEATURES                Location/Qualifiers
     source             1..893
                        /organism="Bos taurus"
                        /organelle="mitochondrion"
                        /strain "Hereford"
                        /db_xref="taxon:9913"
                        /sex="male"
BASE COUNT       2   a    216 c    125 g    260 t
ORIGIN


  Query Match              93.0%;  Score 448.2;  DB 3;  Length 893;
  Best Local Similarity    89.0%;  Pred. No. 1.9e-92;
  Matches  429;  Conservative   51;  Mismatches    1;  Indels    1;  Gaps    1;

Qy        1 gcatata gcaagyacatrayyyctayavyagtacataayrcatayaattattrayygta 60
            |||||||  ||||:||||:|:::|||:|::|||||||||::||||:||||||||:|::||
Db      225 GCATATAA GCAAGTACATGACCTCTATAGCAGTACATAATACATATAATTATTGACTGTA 284

Qy       61 catagta attatrtcaayycatyctyrayaryatatytayyatatayyycytnccayt 120
            |||||||  ||||:||||  ::|||:||::|:|::|||||:||::||||||:::|:| |||:|
Db      285 CATAGTA ATTATGTCAA TTCATTCTTGATAGTATATCTATTATATATTCCTTACCATT 344

Qy      121 agatcac gcttaayta catgccgcgtgaaaccarcaacccgctrrgcagnggatccc 180
            |||||||  ||||||:||  |||||||||||||||||:|||||||||::|||| |||||||
Db      345 AGATCAC AGCTTAATTA CATGCCGCGTGAAACCAGCAACCCGCTAGGCAG-GGATCCC 403

Qy      181 tcttctc tccgggcccatarayygtgggggtcgctatyyartgaaytttaycaggcat 240
            ||||||  |||||||||  ||:|::||||||||||||||::|:||||:||||:|||||||
Db      404 TCTTCTC TCCGGGCCC TAAACCGTGGGGGTCGCTATCCAATGAATTTTACCAGGCAT 463

Qy      241 ctggttc tcttcagggcatctcatctaaaryrgtccattcyttcctcttaaataaga 300
            |||||||  |||||||||  |||||||||||||:::||||||||:||||||||||||||||
Db      464 CTGGTTC TCTTCAGGC CATCTCATCTAAAACGGTCCATTCTTTCCTCTTAAATAAGA 523

Qy      301 catctcgaggactaatgctaatcagcccatgctcacacataactgtgytgtcatacat 360
            |||||||: |||||||||:|||||||||||||||||||||||||||||:||||||||||
Db      524 CATCTCGA GGACTAATC CTAATCAGCCCATGCTCACACATAACTGTGCTGTCATACAT 583
```

```
Qy    361 ttggtatt ttttattttggggatgcttggactcagctatggccgtcaaaggccctgac 420
          |||||||: ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    584 TTGGTATT TTTTATTTTGGGGATGCTTGGACTCAGCTATGGCCGTCAAAGGCCCTGAC 643

Qy    421 ccggagca ctattgtagctggacttaactgcatcttgagcaccagcataatgatargcr 480
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||:||:
Db    644 CCGGAGCA CTATTGTAGCTGGACTTAACTGCATCTTGAGCACCAGCATAATGATAAGCA 703

Qy    481 tg 482
          ||
Db    704 TG 705
```

S1D8

→ 9 hits on this seq in this paper

Also fr. SID 15 (#31 in [illegible] list)

16 #32 [illegible]

```
RESULT    6
BOVWFUA
LOCUS       BOVWFUA       892 bp    DNA             MAM       25-NOV-1994
DEFINITION  Bos indicus (White Fulani strain) DNA fragment.
ACCESSION   L27720
VERSION     L27720.1  GI:443758
KEYWORDS    .
SOURCE      Mitochondrion Bos indicus (strain White Fulani, organelle
            Mitochondrion Bos indicus) female DNA.
  ORGANISM  Mitochondrion Bos primigenius indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Loftus,R.T., MacHugh,D.E., Bradley,D.G., Sharp,P.M. and
            Cunningham,P.
  TITLE     Evidence for two independent domestications of cattle
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (7), 2757-2761 (1994)
  MEDLINE   94195822
FEATURES             Location/Qualifiers
     source          1..892
                     /organism="Bos primigenius indicus"
                     /organelle="mitochondrion"
                     /strain="White Fulani"
                     /db_xref="taxon:9915"
                     /sex="female"
BASE COUNT      291 a    216 c    126 g    259 t
ORIGIN

  Query Match              93.0%;  Score 448.2;  DB 3;  Length 892;
  Best Local Similarity    89.0%;  Pred. No. 1.9e-92;
  Matches  429;  Conservative   51;  Mismatches    1;  Indels    1;  Gaps    1;

Qy        1 gcatataagcaagyacatrayyyctayavyagtacataayrcatayaattattrayygta 60
            |||||||||||||:||||:|:::|||:|::|||||||||::||||:|||||||:|::|||
Db      225 GCATATAAGCAAGTACATGACTTCTATAGCAGTACATAATACATATAATTATTGACTGTA 284

Qy       61 catagtacattatrtcaaayycatyctyrayaryatatytayyatatayyycytnccayt 120
            |||||||||||||:|||||::|||:||::|:|::||||:||::|||||:::|:| |||:|
Db      285 CATAGTACATTATGTCAAATTCATGCTTGATAGTATATCTATTATATATTCCTTACCATT 344

Qy      121 agatcacgagcttaaytaccatgccgcgtgaaaccarcaacccgctrrgcagnggatccc 180
            |||||||||||||||:||||||||||||||||||||:|||||||||::|||| |||||||
Db      345 AGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTAGGCAG-GGATCCC 403

Qy      181 tcttctcgctccgggcccatarayygtgggggtcgctatyyartgaaytttaycaggcat 240
            |||||||||||||||||||||:|::||||||||||||||::|:||||:||||:|||||||
Db      404 TCTTCTCGCTCCGGGCCCATAAACCGTGGGGGTCGCTATCCAATGAACTTTACCAGGCAT 463

Qy      241 ctggttctttcttcagggccatctcatctaaaryrgtccattcyttcctcttaaataaga 300
            ||||||||||||||||||||||||||||||||:::||||||||:||||||||||||||||
Db      464 CTGGTTCTTTCTTCAGGGCCATCTCATCTAAAACGGTCCATTCTTTCCTCTTAAATAAGA 523

Qy      301 catctcgatggactaatgrctaatcagcccatgctcacacataactgtgytgtcatacat 360
            ||||||||||||||||||:||||||||||||||||||||||||||||||:||||||||||
Db      524 CATCTCGATGGACTAATGGCTAATCAGCCCATGCTCACACATAACTGTGCTGTCATACAT 583
```

This result is consistent with the predictions from DNA-bHLH protein co-crystals (Ferr-D'Amar, A. R. et al. (1994) *EMBO J.* 13:180–189; Ferr-D'Amar, A. R. et al. (1993) *Nature* 363:38–45). However, it is noteworthy in that CR8 is the first bHLH vertebrate protein without a leucine zipper (LZ) motif found to bind Class B sites. Protein dimerization is more selective than DNA binding, but currently no rules are available that predict the dimerization preference of any given HLH proteins. Even so, a Class A-binding protein seems to form DNA binding heterodimers only with other Class A proteins, and a bHLH protein with a LZ does not form heterodimers with those without LZs (Blackwood, E. M. et al. (1991) *Science* 251:1211–1217; Prendergast, G. C. et al. (1991) *Cell* 65:395–407). Therefore, if CR8 does form heterodimers, the most likely partner is a class B-binding bHLH protein without a LZ.

Although CR8 is most homologous to hairy in its bHLH domain, its preference for Class B E-box binding sites rather than class C sites, and its lack of a C-terminal WPRW motif, clearly sets CR8 apart and does not predict necessarily that CR8 may act as a transcriptional repressor as do hairy-related proteins. Recently, Id proteins that lack a basic region have been shown to favor proliferation, presumably by forming heterodimers with differentiation inducing bHLH proteins, thereby preventing DNA binding and transcriptional activation of genes that program differentiation (Barone, M. V. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:4985–4988; Hara, E. et al. (1994) *J. Biol. Chem.* 269:2139–2145; Iavarone, A. et al. (1994) *Genes & Dev* 1270–1284). Therefore, CR8 could promote proliferation by suppressing differentiation by either of these transcriptional repressor mechanisms. Alternatively, CR8 could also activate transcription like the bHLH-LZ Myc family.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 35

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2406 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 116..722

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AACCCAACCG CAGTTGACTA GCACCTGCTA CCGCGCCTTT GCTTCCTGGC GCACGCGGAG      60

CCTCCTGGAG CCTGCCACCA TCCTGCCTAC TACGTGCTGC CCTCGCCCCG CAGCC ATG      118

TGC CGC ACC CTG GCC GCC TTC CCC ACC ACC TGC CTG GAG AGA GCC AAA      166

GAG TTC AAG ACA CGT CTG GGG ATC TTT CTT CAC AAA TCA GAG CTG GGC      214

TGC GAT ACT GGG AGT ACT GGC AAG TTC GAG TGG GGC AGT AAA CAC AGC      262

AAA GAG AAT AGA AAC TTC TCA GAA GAT GTG CTG GGG TGG AGA GAG TCG      310

TTC GAC CTG CTG CTG AGC AGT AAA AAT GGA GTG GCT GCC TTC CAC GCT      358

TTC CTG AAG ACA GAG TTC AGT GAG GAG AAC CTG GAG TTC TGG CTG GCC      406

TGT GAG GAG TTC AAG AAG ATC CGA TCA GCT ACC AAG CTG GCC TCC AGG      454

GCA CAC CAG ATC TTT GAG GAG TTC ATT TGC AGT GAG GCC CCT AAA GAG      502

GTC AAC ATT GAC CAT GAG ACC CGC GAG CTG ACG AGG ATG AAC CTG CAG      550

ACT GCC ACA GCC ACA TGC TTT GAT GCG GCT CAG GGG AAG ACA CGT ACC      598

CTG ATG GAG AAG GAC TCC TAC CCA CGC TTC CTG AAG TCG CCT GCT TAC      646

CGG GAC CTG GCT GCC CAA GCC TCA GCC GCC TCT GCC ACT CTG TCC AGC      694

TGC AGC CTG GAC CAG CCC TCA CAC ACC T GAGTCTCCAC GGCAGTGAGG          742
```

```
Qy     361 ttggtatttttttattttgggggatgcttggactcagctatggccgtcaaaggccctgac 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     584 TTGGTATTTTTTTATTTTGGGGGATGCTTGGACTCAGCTATGGCCGTCAAAGGCCCTGAC 643

Qy     421 ccggagcatctattgtagctggacttaactgcatcttgagcaccagcataatgatargcr 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||:||:
Db     644 CCGGAGCATCTATTGTAGCTGGACTTAACTGCATCTTGAGCACCAGCATAATGATAAGCG 703

Qy     481 tg 482
           ||
Db     704 TG 705
```

such as MyoD and AS-C gene products show a cell-type specific expression, others such as E12/E47 and da are fairly ubiquitously expressed. The tissue distribution of CR8 was analyzed using a Multiple Tissue Northern blot. CR8 transcripts of the expected size (3.2 kb) were detected in all tissues examined except placenta (see FIG. 16). FIG. 16 demonstrates that the Multiple Tissue Northern Blot membranes (Clontech; each lane contains 2 fg poly(A)+ RNA from indicated human tissue) were hybridized with human CR8 probe.

The expression of CR8 in peripheral blood leukocytes was unexpected, in that CR8 is not expressed by quiescent T cells. This may reflect much higher sensitivity of Multiple Tissue Northern blot prepared from poly(A)+ RNA compared to our previous Northern blots, which used total RNA. Alternatively, the contribution of other leukocytes such as B cells, NK cells, monocytes and granulocytes that were not present in the original T cell preparations could account for CR8 expression by the peripheral blood leukocytes.

DNA-binding activity of CR8:

The canonical bHLH binding sequence is called the E-box, CANNTG, originally identified in the immunoglobulin heavy chain enhancer (Ephrussi, A. et al. (1985) *Science* 227:134–140). Many bHLH proteins were later divided into two mutually exclusive classes, depending on whether they bind to the Class A sites (CAGCTG/CACCTG) or the Class B sites (CACGTG/CATGTG) (Dang, C. V. et al. (1992) *Proc. Natl. Acad. Sci.* USA 89:599602). The presence of an arginine residue at position 13 ("R13", see FIG. 15A) in the basic region, which CR8 contains, is considered to be the key structural criterion that defines Class B binding specificity. However, despite the presence of "R13", hairy-related bHLH proteins are reported to prefer noncanonical binding sites such as the N-box (CACNAG) (Akazawa, C. et al. (1992) *J. Biol. Chem.* 267:21879–21885; Sasai, Y. et al. (1992) *Genes & Dev* 6:2620–2634; Tietze, K. et al. (1992) *Proc. Natl. Acad. Sci.* USA 89:6152–6156) or the Class C (CACGCG) sites (Ohsako, S. et al (1994) *Genes & Dev* 8:2743–2755; Van Doren, M. et al. (1994) *Genes & Dev* 8:2729–2742.). Therefore, the binding of CR8 to all of these sites was tested.

Since it is well documented that the bHLH domain is sufficient to determine its DNA binding specificity (Pognonec, P. et al. (1994) *Mol. Cell. Biol.* 11:5125–5136), the bHLH domain of CR8 (CR8 bHLH) expressed in *E. coli* was employed for this study. A histidine-tag was added to facilitate the purification of the recombinant protein. While most of the recombinant protein localized in inclusion bodies, there was still enough soluble protein in the cytoplasm, thereby enabling its purification under native conditions using a Ni^{2+} column. A single band of protein was detected at the expected size (16.6 kD with the histidine-tag) by silver staining. EMSA was carried out using this recombinant protein.

FIG. 17A is an EMSA shows binding of recombinant bHLH proteins to the radiolabeled probes. CR8 bHLH protein strongly binds to the Class B (CACGTG, lane 3) and the Class C (CACGCG, lane 4) sites, and weakly to the N box (CACNAG, lane 5) sequence but not to the Class A (CACCTG, lane 2) site. Binding of the bHLH region of da protein to the Class A site is shown as control (lane 1). As shown in FIG. 17A, CR8 bHLH bound to the Class B and the Class C probes, but only weakly to the N-box probe, and not at all to the Class A probe. However, the control da bHLH protein effectively recognized and bound to the same Class A probe.

To examine the relative binding affinity, a large excess of non-labeled oligonucleotide was added to the reaction as competitor. FIG. 17B shows competition of the binding of CR8 bHLH to the Class B sites. 0.5 ng of the radiolabeled Class B probe was incubated with CR8 bHLH in the absence (lane 1) or the presence (lanes 2 to 7) of either 25 ng (50-fold excess; lanes 2, 4 and 6) and 250 ng (500-fold excess; lanes 3, 5 and 7) of unlabeled competitors. FIG. 17B demonstrates that the binding of CR8 bHLH to the radiolabeled Class B site can be abolished partially by a 50-fold excess, and completely by a 500-fold excess of Class B site (lanes 2 and 3), while a 500-fold excess of Class C site only partially displaced CR8 bHLH from the labeled Class B probe (lanes 4 and 5) and the N-box sequence did not affect the binding at all (lanes 6 and 7). Thus, since all these experiments were done in the absence of other HLH proteins, it appears that CR8 bHLH bound to the Class B sequence as a homodimer with the highest affinity.

The CR8 gene encodes a novel bHLH protein that appears to fit into a class by itself. Other than c-myc, CR8 is the first bHLH-containing protein found to be induced by cytokines. Also, from its predicted amino acid sequence, CR8 clearly contains a bHLH motif most closely related to the hairy family, but the amino acid sequence of the basic region differs from other hairy-related proteins: the position of the proline residue is N-terminal to the defining proline of the hairy-related proteins, and CR8 lacks the C-terminal WRPW sequence found in all other hairy-related-related proteins. These differences in the amino acid sequence, especially of the basic region, most likely account for the unique binding specificity of the CR8 bHLH domain. Instead of preferring Class C sites according to the other hairy-related family members (Ohsako, S. et al. (1994) *Genes & Dev* 8:2743–275572; Van Doren, M. et al. (1994) *Genes & Dev* 8:2729–2742), CR8 binds preferentially to Class B sites.

The identification of CR8 as a bHLH protein, thereby functioning, most likely, as a regulator of subsequent gene expression stimulated by IL-2, provides a link between the immediate biochemical events triggered by cytokine receptors and the subsequent events of proliferation and/or differentiation. Thus far, IL-2 has been found to activate the serine/threonine kinase proto-oncogene Raf-1 (Turner, B. et al. (1991) Proc. Natl. Acad. Sci. USA 88:1227–1231; Zmuidzinas, A. et al. (1991) Mol. Cell Biol. 11:2794–2803) and the tyrosine-specific kinases JAK 1 and JAK 3 (Beadling, C. et al. (1994) EMBO J. 13:5605–5615; Miyazaki, T. et al. (1994) *Science* 266:1045–1047; Russell, S. M. et al. (1994) *Science* 266:1042–1045).

From the results described herein comparing the effects of IFNa and forskolin on CR8 and c-myc gene expression, the regulation of these two bHLH genes is clearly distinct. It is also of interest that although IFNβ antagonizes IL-2-promoted cell cycle progression, it promotes the expression of both CR8 and c-myc. Indeed, induction of c-myc by IFNβ was unexpected, as it was previously reported to be suppressed by IFNs (Einat, M. et al. (1985) *Nature* 313:597–600). The bHLH region of CR8 is most homologous to that of hairy and the bHLH proteins of the E(spl)-C. In Drosophila, the hairy-related bHLH proteins function as transcriptional repressors, and this activity requires the basic DNA binding region, as well as the interaction with a non-HLH protein termed groucho (gro) via the C-terminal WRPW motif (Paroush, Z. et al. (1994) *Cell* 79:805–815). Although mammalian homologues of gro have been identified (Stifani, S. et al. (1992) *Nat. Genet.* 2:119–127), they are not likely to interact with CR8 because CR8 lacks the WRPW motif.

The results described herein indicate that CR8 recombinant protein binds to Class B E-box sites as a homodimer.

1

```
RESULT    4
BOVHERB/c
LOCUS          BOVHERB        893 bp     DNA               MAM        25-NOV-1994
DEFINITION     Bos taurus (Hereford strain) DNA fragment.
ACCESSION      L27725
VERSION        L27725.1   GI:443748
KEYWORDS       .
SOURCE         Mitochondrion Bos taurus (strain Hereford, organelle Mitochondrion
               Bos taurus) male DNA.
  ORGANISM     Mitochondrion Bos taurus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
               Bovidae; Bovinae; Bos.
REFERENCE      1  (bases 1 to 893)
  AUTHORS      Loftus,R.T., MacHugh,D.E., Bradley,D.G., Sharp,P.M. and
               Cunningham,P.
  TITLE        Evidence for two independent domestications of cattle
  JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 91 (7), 2757-2761 (1994)
  MEDLINE      94195822
FEATURES                Location/Qualifiers
     source             1..893
                        /organism="Bos taurus"
                        /organelle="mitochondrion"
                        /strain="Hereford"
                        /db_xref="taxon:9913"
                        /sex="male"
BASE COUNT       292 a     216 c     125 g     260 t
ORIGIN


  Query Match              100.0%;   Score 43;   DB 3;   Length 893;
  Best Local Similarity    100.0%;   Pred. No. 6.3e-06;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 taatgtccatgcttatcattatgctggtgctcaagatgcagtt 43
            |||||||||||||||||||||||||||||||||||||||||||
Db      712 TAATGTCCATGCTTATCATTATGCTGGTGCTCAAGATGCAGTT 670
```

[illegible]

filing date – [illegible]

-continued

TAC CTG CAG CTG TAC GCC ACC CTG CTG CTG CTT CTC ATT GTC TCG GTG 861

CTC GCC TGC AAC TTC AGT GTC ATT CTC AAC CTC ATC CGC ATG CAC CGC 909

CGA AGC CGG AGA AGC CGC TGC GGA CCT TCC CTG GGC AGT GGC CGG GGC 957

GGC CCC GGG GCC CGC AGG AGA GGG GAA AGG GTG TCC ATG GCG GAG GAG 1005

ACG GAC CAC CTC ATT CTC CTG GCT ATC ATG ACC ATC ACC TTC GCC GTC 1053

TGC TCC TTG CCT TTC ACG ATT TTT GCA TAT ATG AAT GAA ACC TCT TCC 1101

CGA AAG GAA AAA TGG GAC CTC CAA GCT CTT AGG TTT TTA TCA ATT AAT 1149

TCA ATA ATT GAC CCT TGG GTC TTT GCC ATC CTT AGG CCT CCT GTT CTG 1197

AGA CTA ATG CGT TCA GTC CTC TGT TGT CGG ATT TCA TTA AGA ACA CAA 1245

GAT GCA ACA CAA ACT TCC TGT TCT ACA CAG TCA GAT GCC AGT AAA CAG 1293

GCT GAC CTT T GAGGTCAGTA GTTTAAAAGT TCTTAGTTAT ATAGCATCTG 1343

GAAGATCATT TTGAAATTGT TCCTTGGAGA AATGAAAACA GTGTGTAAAC AAAATGAAGC 1403

TGCCCTAATA AAAAGGAGTA TACAAACATT TAAGCTGTGG TCAAGGCTAC AGATGTGCTG 1463

ACAAGGCACT TCATGTAAAG TGTCAGAAGG AGCTACAAAA CCTACCCTCA GTGAGCATGG 1523

TACTTGGCCT TTGGAGGAAC AATCGGCTGC ATTGAAGATC CAGCTGCCTA TTGATTTAAG 1583

CTTTCCTGTT GAATGACAAA GTATGTGGTT TTGTAATTTG TTTGAAACCC CAAACAGTGA 1643

CTGTACTTTC TATTTTAATC TTGCTACTAC CGTTATACAC ATATAGTGTA CAGCCAGACC 1703

AGATTAAACT TCATATGTAA TCTCTAGGAA GTCAATATGT GGAAGCAACC AAGCCTGCTG 1763

TCTTGTGATC ACTTAGCGAA CCCTTTATTT GAACAATGAA GTTGAAAATC ATAGGCACCT 1823

TTTACTGTGA TGTTTGTGTA TGTGGGAGTA CTCTCATCAC TACAGTATTA CTCTTACAAG 1883

AGTGGACTCA GTGGGTTAAC ATCAGTTTTG TTTACTCATC CTCCAGGAAC TGCAGGTCAA 1943

GTTGTCAGGT TATTTATTTT ATAATGTCCA TATGCTAATA GTGATCAAGA AGACTTTAGG 2003

AATGGTTCTC TCAACAAGAA ATAATAGAAA TGTCTCAAGG CAGTTAATTC TCATTAATAC 2063

TCTTTATCCT ATTTCTGGGG GAGGATGTAC GTGGCCATGT ATGAAGCCAA ATATTAGGCT 2123

TAAAAACTGA AAAATCTGGT TCATTCTTCA GATATACTGG AACCCTTTTA AAGTTGATAT 2183

TGGGGCCATG AGTAAAATAG ATTTTATAAG ATGACTGTGT TGTACTAAAA TTCATCTGTC 2243

TATATTTTAT TTAGGGGACA TGGTTTGACT CATCTTATAT GGGAAACCAT GTAGCAGTGA 2303

GTCATATCTT AATATATTTC TAAATGTTTG GCATGTAAAC GTAAACTCAG CATCACAATA 2363

TTTCAGTGAA TTTGCACTGT TTAATCATAG TTACTGTGTA AACTCATCTG AAATGTTACC 2423

AAAAATAAAC TATAAAACAA AATTTGA 2450

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 358 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gly Asn Ala Ser Asn Asp Ser Gln Ser Glu Asp Cys Glu Thr Arg
1 5 10 15

Gln Trp Phe Pro Pro Gly Glu Ser Pro Ala Ile Ser Ser Val Met Phe
20 25 30

Ser Ala Gly Val Leu Gly Asn Leu Ile Glu Leu Ala Leu Leu Ala Arg